DOCUMENT RESUME

ED 367 957 CS 011 635

TITLE	Standards for the Assessment of Reading and Writing.
INSTITUTION	International Reading Association, Newark, Del.; National Council of Teachers of English, Urbana, Ill.
REPORT NO	ISBN-0-8141-0213-1; ISBN-0-87207-674-1
PUB DATE	94
NOTE	47p.; Prepared by the IRA/NCTE Joint Task Force on Assessment.
AVAILABLE FROM	Order Department, International Reading Association, 800 Barksdale Road, P.O. Box 8139, Newark, DE 19714-8139 (Book No. 674; $4.95 members, $6.95 nonmembers).
PUB TYPE	Guides - Non-Classroom Use (055)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Academic Standards; Elementary Secondary Education; Instructional Improvement; Parent Participation; *Reading Achievement; *Reading Instruction; *Student Evaluation; *Writing Achievement; *Writing Instruction
IDENTIFIERS	*Educational Issues

ABSTRACT

Reflecting advances in the understanding of language, learning, and the complex literacy demands of a rapidly changing democratic and technological society, this booklet provides a set of standards to guide decisions about assessing the teaching and learning of reading and writing. After an introduction that presents an overview of what is meant by assessment, the booklet discusses the following 11 standards: (1) the interests of the student are paramount in assessment; (2) the primary purpose of assessment is to improve teaching and learning; (3) assessment must reflect and allow for critical inquiry into curriculum and instruction; (4) assessment must recognize and reflect the intellectually and socially complex nature of reading and writing and the important roles of school, home, and society in literacy development; (5) assessment must be fair and equitable; (6) the consequences of an assessment procedure are the first, and most important, consideration in establishing the validity of the assessment; (7) the teacher is the most important agent of assessment; (8) the assessment process should involve multiple perspectives and sources of data; (9) assessment must be based in the school community; (10) all members of the educational community--students, parents, teachers, administrators, policy makers, and the public--must have a voice in the development, interpretation, and reporting of assessment; and (11) parents must be involved as active, essential participants in the assessment process. The booklet offers rationales and implications for each standard, and some standards are accompanied with case studies. A glossary of assessment terminology is attached. (RS)

ED 367 957

Standards for the Assessment of Reading and Writing

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

L. W. Kline

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

International Reading Association *and*
National Council of Teachers of English

Prepared by the IRA/NCTE Joint Task Force on Assessment

CS 011635

BEST COPY AVAILABLE

Standards for the Assessment of Reading and Writing

International Reading Association *and*
National Council of Teachers of English

Prepared by the IRA/NCTE Joint Task Force on Assessment
Peter Johnston, Chair, SUNY at Albany
Johnny Tolliver, Co-Chair, Delaware State University
Carl Braun, University of Calgary
James E. Davis, Ohio University
Alan Farstrup, International Reading Association
Shirley Haley-James, Georgia State University
Jerome C. Harste, Indiana University
Marjorie Y. Lipson, University of Vermont
Marianne Marino, Central School, NJ
Miles Myers, National Council of Teachers of English
P. David Pearson, University of Illinois at Urbana-Champaign
John J. Pikulski, University of Delaware
Susan Stires, Center for Teaching and Learning, ME
Jackie E. Swensson, Metropolitan State College, CO
Judith N. Thelen, Frostburg State University, MD
Sheila D. Valencia, University of Washington

The work of the Joint Task Force on Assessment of the International Reading Association (IRA) and the National Council of Teachers of English (NCTE) and the publication of this document were funded in part by the John D. and Catherine T. MacArthur Foundation. The IRA/NCTE Joint Task Force was formed in 1991 by the IRA Executive Board and the NCTE Executive Committee. *Standards for the Assessment of Reading and Writing* is the report of this Joint Task Force. The IRA Board of Directors and the NCTE Executive Committee approved *Standards for the Assessment of Reading and Writing* in October 1993. Members of these boards are listed below.

IRA Board of Directors
Doris D. Roettger (President), *Heartland Area Education Agency, Johnston, IA;* Susan Mandel Glazer (President-elect), *Rider College, NJ;* Dolores B. Malcolm (Vice President), *St. Louis, MO, Public Schools;* Mabel T. Edmonds, *St. Louis, MO, Public Schools;* Linda B. Gambrell, *University of Maryland-College Park;* Daniel R. Hittleman, *Queens College, City University of New York;* Jerry L. Johns, *Northern Illinois University;* Sandra McCormick, *Ohio State University;* Donna M. Ogle, *National-Louis University;* Pehr-Olof Rönnholm, *Cygnaeus School, Turku, Finland;* MaryEllen Vogt, *California State University, Long Beach;* Carmelita Kimber Williams, *Norfolk State University;* Alan E. Farstrup, *Executive Director, IRA.*

NCTE Executive Committee
Jesse Perry (President), *San Diego City Schools,* ret.; Janie Hydrick (President-elect), *Mesa Public Schools, AZ;* Miriam T. Chaplin (Vice President), *Rutgers University;* James E. Davis (Past President), *Ohio University;* Marianne Marino (Representative at Large), *Central School, Glen Rock, NJ;* Connie Sears (Representative at Large), *Weatherford High School, OK;* Barbara Flores (Elementary Section Chair), *California State University;* Mildred Miller (Secondary Section Chair), *Pinacate Middle School, Perris, CA;* William G. McBride (Secondary Section Associate Chair), *Colorado State University;* Cynthia Selfe (College Section Chair), *Michigan Technological University;* Anne Ruggles Gere (CCCC Chair), *University of Michigan;* Joseph Milner (CEE Chair), *Wake Forest University;* Donald Stephan (CEL Chair), *Sidney High School, OH.* Headquarters staff who are members of the Executive Committee are Miles Myers (Executive Director), Charles Suhor (Deputy Executive Director), Karen Smith (Associate Executive Director), and Katherine Hope (Associate Executive Director for Business).

Design: R. Maul

IRA Stock Number: 674

NCTE Stock Number: 02131

© 1994 by the International Reading Association and the National Council of Teachers of English. Published by the International Reading Association and the National Council of Teachers of English. All rights reserved. Printed in the United States of America.

It is the policy of NCTE in its journals and other publications to provide a forum for the open discussion of ideas concerning the content and the teaching of English and the language arts. Publicity accorded to any particular point of view does not imply endorsement by the Executive Committee, the Board of Directors, or the membership at large, except in announcements of policy, where such endorsement is clearly specified.

IRA ISBN: 0-87207-674-1

NCTE ISBN: 0-8141-0213-1

Contents

Introduction 5
- The Nature of Assessment 5
- The Nature of Language 7
- The Learning of Language 9
- The Assessment of Language 10
- The Language of Assessment 11

The Standards 13

Goals

1. The interests of the student are paramount in assessment. 13
   - Rationale 13
   - Implications 14
2. The primary purpose of assessment is to improve teaching and learning. 15
   - Rationale 15
   - Implications 16
3. Assessment must reflect and allow for critical inquiry into curriculum and instruction. 17
   - Rationale 17
   - Implications 18
   - Case Study: Instruction Informs Assessment 18
4. Assessments must recognize and reflect the intellectually and socially complex nature of reading and writing and the important roles of school, home, and society in literacy development. 19
   - Rationale 19
   - Implications 20
   - Case Study: Assessment Must Be Contextualized 21
5. Assessment must be fair and equitable. 21
   - Rationale 21
   - Implications 24
6. The consequences of an assessment procedure are the first, and most important, consideration in establishing the validity of the assessment. 25
   - Rationale 25
   - Implications 26

Implementation

7. The teacher is the most important agent of assessment. 27
   - Rationale 27
   - Implications 28

8. The assessment process should involve multiple perspectives and sources of data. 29
Rationale 29
Implications 30
Case Study: Special Education, People Instead of Tests, and John 31
9. Assessment must be based in the school community. 33
Rationale 33
Implications 33
Case Study: All Stakeholders Must Participate in Assessment 34
10. All members of the educational community—students, parents, teachers, administrators, policy makers, and the public—must have a voice in the development, interpretation, and reporting of assessment. 35
Rationale 35
Implications 36
11. Parents must be involved as active, essential participants in the assessment process. 37
Rationale 37
Implications 38

Appendix: Glossary of Assessment Terminology 39

Introduction

This document provides a set of standards to guide decisions about assessing the teaching and learning of reading and writing. In the past thirty years, research has produced revolutionary changes in our understanding of language, learning, and the complex literacy demands of a rapidly changing democratic and technological society. The standards proposed in this document are intended to reflect these advances in our understanding.

Most people reading this document share common school experiences with respect to literacy and assessment. For example, in school we read to get the correct meaning of a text so that we could answer questions put to us by someone who already knew the correct meaning—or by a test (often multiple-choice) for which the correct answers were already determined. In order to develop assessment practices that serve students in an increasingly complex society, we must outgrow the limitations of our own schooling histories and understand language, literacy, and assessment in more complex ways. Literacy, for example, involves not just reading and writing, but a wide range of related language activities. It is both more social and more personal than a mere set of skills.

To improve language assessment we must understand not only assessment, but language and how it relates to assessment. The need to understand language is particularly important. Language is not only the thing being assessed, but also part of the process of assessment itself. Consequently, any discussion of literacy assessment must include a discussion of language: what it is, how it is learned, and how it relates to assessment. Before we state our assessment standards, then, we will give an overview of what we mean by assessment, and how we understand language and its relationship to assessment.

The Nature of Assessment

At one point in the history of schooling, a transmission view of knowledge, curriculum, and assessment dominated (or appeared to satisfy) our social, political, and economic needs. Knowledge was regarded as a static entity that was "out there" somewhere, so the key educational question was, "How do you get it from out there into the students' heads?" Then the corollary assessment question was, "What counts as evidence that the knowledge really is in their heads?" In a transmission view, it made sense

to develop educational standards that specified the content of instruction before developing assessment procedures and engagements.

In our postindustrial society, with its ever-increasing need for workers with strong problem-solving dispositions and skills, one of the basic purposes of schooling that has gained in prominence since the 1920s has become not simply the transmission of knowledge but the more complex nurturing of independent learning. A curriculum committed to independent learning is built upon the premise that inquiry, rather than mere transmission of knowledge, is the basis of teaching and learning.

This shift from knowledge transmission to inquiry as a primary goal of schools has important implications for assessment. In a knowledge-transmission framework, tests of static knowledge can suffice as assessment instruments. Students are primarily accountable (either they got the knowledge or they didn't), with teachers held accountable next. Policy makers, such as school board members, trustees, or regents are then the primary recipients of assessment data.

Inquiry changes the role of assessment and the roles of the participants. Within an inquiry framework, assessment is the exploration of how the educational environment and the participants in the educational community support education as a process of learning to become independent thinkers and problem solvers. This exploration includes an examination of the environment for teaching and learning, the processes and products of learning, and the degree to which all participants (students, teachers, administrators, parents, and board members) meet their obligation to support inquiry.

Inquiry emphasizes different processes and types of knowledge than does knowledge transmission. For example, it values the ability to recognize problems and to generate multiple and diverse perspectives in trying to solve them. An inquiry perspective asserts that while knowledge and language are likely to change over time, what will remain constant is the need for learners at all levels (students, teachers, parents, administrators, and policy makers) to solve new problems, generate new knowledge, and invent new language practices. An inquiry perspective would promote problem finding and problem solving as goals for all participants in the educational community. For example, inquiry would value the question of how information from different sources can be used to solve a particular problem. It would value explorations of how teachers can promote critical thinking for all students. And it would raise the question of why our society privileges the knowledge and cultural heritage of some groups over others within current school settings.

Inquiry fits the needs of a multicultural society in which it is essential to value and find strength in cultural and epistemological diversity. It also honors the commitment to raising questions and generating multiple so-

lutions. Various stakeholders and cultural groups provide different answers and new perspectives on problems. In a sense, respecting difference among learners becomes a self-correcting and enriching strategy for all learners and, most important, for the curriculum.

Just as the principle of inquiry values differences, so the principle of difference values conversation over recitation as the primary mode of discourse. In a recitation, it is assumed that one person, the teacher, possesses the answers and that the others, the students, interact with the teacher and one another in an attempt to uncover the teacher's knowledge. In a conversation, all of the stakeholders in the educational environment (students, parents, teachers, specialists, administrators, and policy makers) have a voice at the table as curriculum, standards, and assessments are negotiated. Neither inquiry nor learning is viewed as the exclusive domain of students and teachers. Both are primary concerns for all members of the school community. For example, administrators ask themselves hard questions about whether the structures they have established support staff development, teacher reflection, and student learning. School board members ask themselves whether they have lived up to the standards they have set for themselves and their schools to provide teachers and students with the resources they need to guarantee learning opportunities.

Quality assessment, then, hinges on the process of setting up conditions so that the classroom, the school, and the community become centers of inquiry where students, teachers, and other members of the school community investigate their own learning, both individually and collaboratively. The onus of assessment does not fall disproportionately upon students and teachers (which is often the case in schools in our society today); instead, all those involved in curriculum inquiry are held responsible for investigating the roles they have played. Different members of the school community have different but interacting interests, roles, and responsibilities. And assessment is the medium that allows all to explore what they have learned and whether they have met their responsibilities to the school community.

The Nature of Language

Language is very much like a living organism. It cannot be put together from parts like a machine, and it is constantly changing. Like a living organism it exists only in interaction with others, in a social interdependence. Language is a system of signs through and within which we represent and make sense of the world and of ourselves. Language does not contain meaning; rather, meaning lies in the social relationships within which language occurs. Individuals in communities make sense of lan-

guage within their social relationships, their personal histories, and their collective memory. In order to make sense of even a single word, people take into account the situation and their relationship with the speaker or writer, as they understand all three—the word, the situation, and their relationship with speaker or writer.

Take, for example, the word "family," a word used many times as if all members of the society were agreed on its meaning. The word may mean different things in different contexts, however, whether cultural, situational, or personal. To a white forty-year-old whose parents moved with their two children from Illinois to Maine in the fifties, "family" may mean the nuclear family with which she grew up. To someone from a different culture, for example an African American or Asian American, the word "family" may conjure images of the constellation of grandparents, aunts, uncles, and cousins who live together or near each other and constitute his family. In yet another context, for example a television commercial for a long-distance telephone company, appeals to the word "family" take on a persuasive edge that may or may not coincide with the hearer's own needs or wants. And meaning may vary even between one person and another, since the meanings attached to the word "family" are likely to be quite different depending on one's own experience in the family or families one has lived with. Thus different people make different sense of apparently similar language. Even the same words will mean different things to the same person in different situations and to different people in what seems to be the same situation.

Because individuals make different sense of language depending on their cultural and personal histories, there are complexities that we do not often take seriously. For example, when we attempt to standardize a test (make it the same for everyone), we make the tenuous assumption that students will all make the same meaning from the language of our instructions.

Different cultures also have different ways of representing the world, themselves, and their intentions with language. For example, in any given cultural group, people have different ways of greeting one another, depending on the situation (e.g., a business meeting, a funeral, a date) and on their relationship to one another. Our own language practices come from our cultural experience with language, but our individual language practices along with those of others collectively make the culture. Indeed, the different ways people use language to make sense of the world and of their lives are the major distinguishing features of different cultural groups.

At the same time, language is always changing as we use it. Words acquire different meanings, and new language structures and uses appear as people stretch and pull the language to make new meanings. Conse-

quently, the meaning that individuals make from language varies across time, social situation, personal perspective, and cultural group.

The Learning of Language

By the time children arrive at school they have learned to speak at least one language, and have mastered most of the language structures they will ever use. Through social interaction, using the language they hear around them from birth, they have developed vocabulary, syntax, and underlying (but unknown to them) rules of grammar that give meaning to the world as they see it. Nonetheless, we often teach language in schools as if children came with little or no language competence. Nothing could be further from the truth. Children can request, demand, explain, recount, persuade, and express opinions. They bring to school the ability to narrate their own life histories. They are authors creating meaning with language long before they arrive at school.

As children acquire language in social interaction, particularly with others whose language is different or more complex, they gain flexibility, engaging in language for different purposes and in different social situations. Learning a second dialect roughly parallels learning the first, for learning any language also entails becoming competent in the social relationships that underlie it. Children also develop fluent use of the language without explicit knowledge of or instruction in rules and grammars. We would most productively teach grammars and rules as tools we use for analyzing language after it has been acquired, since even most adults who have considerable facility with the language can articulate few, if any, grammar or language rules. In spite of this truism, we often go about assessment and instruction in schools as if this were not the case. Furthermore, although we pretend otherwise, language is not acquired in any simple hierarchical sequence.

In some ways, school actually plays a modest role in language acquisition, the bulk of which occurs outside the school. In schools we must learn to teach language in a way that preserves and respects students' individuality at the same time that we empower them to learn how to be responsible and responsive members of learning communities. In other words, we must respect their right to their own interpretations of language, including the texts they read and hear, but we must help them learn that meaning is negotiated with other members of the learning communities within which they live and work. To participate in that negotiation, they must understand and be able to master the language practices and means of negotiation of the cultures within which they live. They must understand the language conventions that are sanctioned in different social situations, and the consequences of adhering to or violating those conventions.

Although much of our language is learned outside school, studying language is the foundation of schooling, not just of the "Language Arts." For example, in "science" class in school, we make knowledge of the world using language. To study science, then, we must study the language through which we make scientific knowledge, language that has an important impact on the curriculum. For example, if in reading and writing about science the language is dispassionate and distancing, then that is part of the knowledge that students construct about science, part of the way they relate to the world through science.

The Assessment of Language

Our description of language and language learning has important implications for the assessment of language, first because it is the object of assessment—the thing being assessed—and second, because it is the medium of assessment—the means through and within which we assess. Instructional outcomes in the language arts, and assessment policies and practices, should reflect what we know about language and its acquisition. For example, to base a test on the assumption that there is a single correct way to write a persuasive essay is a dubious practice. Persuading someone to buy a house is not the same as persuading someone to go on a date. Persuading someone in a less powerful position is not the same as persuading someone in a more powerful position—which is to say that persuasive practices differ across situations, and purposes, and cultural groups. Similarly, that texts can (and should) be read from different perspectives must be taken as a given—a goal of schooling not to be disrupted by assessment practices that pretend otherwise. To assert through a multiple-choice test that a piece of text has only one meaning is unacceptable given what we know of language.

Moreover, to the extent that assessment practices legitimize only the meanings and language practices of particular cultural groups, these practices are acts of cultural oppression. When our assessments give greater status to one kind of writing over another—for example, expository writing over narrative writing—we are making very powerful controlling statements about the legitimacy of particular ways of representing the world. These statements may tend to be reflected in classroom practices.

When we attempt to document students' language development, we are partly involved in producing that development. For example, if we decide that certain skills are "basic" and some are "higher level," and that the one needs to be acquired before the other, that decision affects the way we organize classrooms, plan our teaching, group students, and discuss reading and writing with them. The way we teach literacy, the way we sequence lessons, the way we group students, even the way we physically arrange the classroom all have an impact on their learning.

The Language of Assessment

Because it involves language, assessment is an interpretive process. Just as we construct meanings for texts that we read and write, so do we construct "readings" or interpretations of our students based upon the many "texts" they provide for us. These assessment texts come in the form of the pieces that students write, their responses to literature, the various assignments and projects they complete, the contributions they make to discussions, their behavior in different settings, the questions they ask in the classroom or in conferences, their performances or demonstrations involving language use, and tests of their language competence. Two different people assessing a student's reading or writing, his or her literate development, may use different words to describe it.

In classrooms, teachers assess students' writing and reading and make evaluative comments about writers whose work is read. The language of this classroom assessment becomes the language of the literate classroom community and thus becomes the language through which students evaluate their own reading and writing. If the language of classroom assessment implies that there are several interpretations of any particular text, students will come to gain confidence as they assess their own interpretations, and will value diversity in the classroom. If, on the other hand, the language of classroom assessment implies that reading and writing can be reduced to a simple continuum of quality, students will assess their own literacy only in terms of their place on that continuum relative to other students, without reflecting productively on their own reading and writing practices.

When teachers write report cards, they are faced with difficult language decisions. They must find words to represent a student's literate development in all its complexity, often within severe time, space, and format constraints. They must also accomplish this within the diverse relationships and cultural backgrounds among the parents, students, and administrators who might read the report. Some teachers are faced with reducing extensive and complex knowledge about each student's development to a single word or letter. This situation confronts them with very difficult ethical dilemmas. Indeed, the greater the knowledge the teacher has of the student's literacy, the more difficult this task becomes.

But it is not just classroom assessment that is interpretive. The public "reads" students, teachers, and schools from the data they are given. Parents make sense of a test score or a report card grade or comment based on their own schooling history, beliefs, and values. A test score may look "scientific" and "objective," but it too must be interpreted, which is always a subjective and value-laden process.

The terms in which people discuss students' literacy development have also changed over time. For example, at various points in our recent his-

tory, students considered to be having difficulty becoming literate have acquired different labels, such as "basic writer," "remedial reader," "disadvantaged," "learning disabled," "underachiever," or "retarded reader." These different terms can have quite different consequences. Students described as "learning disabled" are often treated and taught quite differently from students who are similarly literate but described as "remedial readers."

But assessment itself is the object of much discussion, and the language of that discussion is also important. For example, teachers' observations are often described as "informal" and "subjective" and contrasted with test results that are considered "formal" and "objective." The knowledge constructed in such a discussion would be quite different from that constructed in a discussion in which teachers' observations were described as "direct documentation" and test results as "indirect estimation."

Assessment terms change as different groups appropriate them for different purposes, and as situations change. Recent discussions about assessment have changed some of the ways in which previously reasonably predictable words are used, belying the simplicity of the glossary we include at the end of this document. For example, the term "norm-referenced" once meant that assessment data on one student, typically test data, were interpreted in comparison with the data of other students who were considered similar. A norm-referenced interpretation of a student's writing might assert that it is "as good as that of 20 percent of the students that age in the country." Similarly, "criterion-referenced" assessment once meant simply that a student's performance was interpreted with respect to a particular level of performance—either it met the criterion or it did not. Recently, however, it has become much less clear in which ways these terms are being used. The line between criterion and norm has broken down. For example, "criterion" has recently come to mean "dimension" or "valued characteristic." "Norm" has come to be used in much the same sense. But even in the earlier (and still more common) meaning, most criteria for criterion-related tests are arrived at by finding out how a group of students performs on the test, and then setting criteria in accord with what seems a reasonable point for a student's passing or failing the test.

The Standards

1. The interests of the student are paramount in assessment.

Rationale

This standard refers to individual students, not students on average nor students collectively. Assessment must serve, not harm, each and every student. This means that each individual's intellectual, social, and emotional well-being must be considered, even when the decision to be made from the assessment will affect other individual students or even an entire class or school.

We must recognize that assessment experiences, formal or informal, have consequences for students (see standard 6—consequential validity). Assessment procedures have profound effects on students' lives. Assessments may alter their educational opportunities, increase or decrease their motivation to learn, elicit positive or negative feelings about themselves and others, and influence their understanding of what it means to be literate, educated, or successful.

What features of assessment are likely to serve the student's interests? First and foremost, assessment must encourage students to reflect on their own reading and writing in productive ways, to evaluate their own intellectual growth, and to set goals. In this way, students become involved in and responsible for their own learning and better able to assist the teacher in focusing instruction. Past assessment practices, particularly normative practices, have often produced conditions of threat and defensiveness for students. Constructive reflection is particularly difficult under such conditions. Thus assessment should emphasize what students can do rather than what they cannot do. Portfolio assessment, for example, if managed properly, can be reflective, involving students in their own learning and assisting teachers in refocusing their instruction.

Second, assessment must provide useful information to inform and enable reflection. The information must be both specific and timely. Specific information on student's knowledge, skills, strategies, and attitudes helps teachers, parents, and students set goals and plan instruction more thoughtfully. Information about students' confusions, counterproductive strategies, and limitations, too, can help students and teachers reflect on and learn about students' reading and writing as long as it is provided in the context of clear descriptions of what they *can* do. The

timeliness of the information is equally important. If information from assessment is not provided immediately, it is not likely to be used. Nor is it likely to be useful, because needs, interests, and aspirations are likely to change with the passage of time. In either case the opportunity to influence and promote learning may be missed.

Third, the assessment must yield high-quality information. The quality of information is suspect when tasks are too difficult or too easy, when students do not understand the tasks or cannot follow the directions, or when they are too anxious to be able to do their best or even their typical work. In these situations students cannot produce their best efforts or demonstrate what they know. Requiring students to spend their time and efforts on assessment tasks that do not yield high-quality, useful information results in students' losing valuable learning time. Such a loss does not serve their interests and is thus an invalid practice (see standard 6).

Implications

This standard implies that if any individual student's interests are not served by an assessment practice, regardless of whether it is intended for administration or decision making by an individual or by a group, then that practice is not valid for that student. Since group-administered, machine-scorable tests do not normally encourage students to reflect constructively on their reading and writing, do not provide specific and timely feedback, and generally do not provide high-quality information about students, they seem unlikely to serve the best interests of students. Similarly, many less formal classroom assessments fail to meet these criteria. Regardless of the source or motivation for any particular assessment, states, school districts, schools, and teachers must demonstrate how these assessment practices benefit and do not harm individual students.

Assessment instruments or procedures themselves are not the only consideration in this standard. The context in which they are used can be equally important. For example, portfolio assessment that satisfies this standard when used in one class may also satisfy it in the context of a high-stakes assessment, such as an accountability assessment in which comparative scores are published in the newspaper. Authentic assessment tasks such as those being tested in California and in the New Standards Project in 112 school districts across the nation offer exciting and insightful possibilities for producing useful information. Students will perform "authentic" or "real-life" tasks over time, and these tasks can be evaluated at the district, state, and national levels and provide much more meaningful information about what a student knows and is able to do. Rather than a simple comparative reporting of aggregate test scores by school or district, which provides numbers only and is more likely to produce

defensiveness and anxiety than insight, such task-oriented assessments can produce meaningful information that shows the level of teaching and learning actually taking place in a learning community.

Indeed, the most powerful assessments for students are likely to be those that occur in the daily activity of the classroom. Maximizing the value of these for students and minimizing the likelihood that they are damaging will involve an investment in staff development and the creation of conditions that enable teachers to reflect on their own practice.

2. The primary purpose of assessment is to improve teaching and learning.

Rationale

Assessment is used in educational settings for a variety of purposes, such as diagnosing reading and writing difficulties, determining eligibility, evaluating programs, and reporting to others. Underlying all these purposes is a concern for improving teaching and learning. However, many times this concern is obscured by the format and process of assessment, and by the ways we use assessment information. For example, although special-education assessments to determine eligibility for special services appear, on the face, to be designed to improve learning (since they are intended to provide access to more appropriate instruction), the improved learning may or may not result.

Currently, literacy assessment is often used for certification and gatekeeping purposes. In that regard, it is common to argue for "high standards." However well-meaning, the result of these practices is to deny access to appropriate instructional opportunities to both individuals and entire groups. It is our belief that no society can afford a selective approach to literacy learning. Indeed, we believe that the need for higher standards is better expressed as the need for higher-quality instruction, for without it "higher standards" simply means denying greater numbers of students access to programs and opportunities. The central function of assessment, therefore, is not to prove whether or not teaching or learning have taken place, but to improve the quality of teaching and learning, and to increase the likelihood that all members of society will acquire a full and critical literacy (see standard 1).

No matter how elaborate and precise are the data provided by an assessment procedure, its interpretation, its use, or the context of its use can render it useless, or worse, with respect to instructional improvement. Consequently, ensuring that assessment leads to instructional improvement is not simply a technical matter of devising instruments for gener-

ating higher-quality data. At least as important are the conditions under which assessment takes place and the climate produced by assessment practices. For example, climates in which assessment data are used to place blame lead to defensiveness rather than to learning and problem solving. On the other hand, even assessment procedures that appear to provide a relatively limited picture of literate learning can provide productive grounds for reflection in the right circumstances.

The closer the assessment process gets to the student and to the context in which he or she is learning, the more likely it is that this primary purpose will be well served. The further away one gets, the more abstract the data become, the less relevant they seem, and the more difficult it becomes to use such data to improve teaching and learning. Both physical and psychological distance are relevant in this context. For example, the larger the population of the classroom, school, or school district, the more difficult it is to maintain psychological closeness to the individual student or teacher.

Keeping assessment consistent with this standard requires teachers, schools, and school districts not only to use assessment to reflect on learning, but to examine, constantly and critically, the assessment process itself.

Implications

Improvement of instruction through assessment cannot be accomplished by imposing external demands. The most productive way for learners to make change is for them to see, for themselves, any conflicts arising in their own actions, understandings, and intentions. Reflectiveness is not just an add-on to instruction in reading and writing; it is an essential component of education. Students who have not learned to reflect on their own learning, who must depend on others to know how their learning is progressing or which learning strategies are working, are not well prepared for survival in a democratic society saturated with choices and complex decisions. Neither will they contribute to the survival of such a society in an age of constantly changing knowledge and conditions.

The assessment problem becomes, then, one of setting conditions so that the classroom becomes a center of inquiry where students and teacher investigate their own learning, both individually and as a learning community.

3. Assessment must reflect and allow for critical inquiry into curriculum and instruction.

Rationale

We asserted at the beginning of this document that assessment is best viewed as curriculum inquiry. In order for inquiry to proceed, the assessment must reflect the full complexity of the curriculum. This is important at all levels of assessment. Policy makers no less than teachers and students must have a clear perception of the curriculum in order to make informed decisions. Decisions based on severely restricted or distorted perceptions will be poor decisions.

In the past, two major problems have beset efforts to align curriculum and assessment. The first has been that curriculum has reflected assessment rather than the other way around. Thus we have often achieved a trivial curriculum by aligning it with trivial assessments. Second, in general, the higher the stakes of the assessment procedure, the more narrow and trivial has been the assessment instrument used. Such assessment practices are common even when the consequences of these assessments are very serious for individual lives, such as promotion, or placement, or public accountability. For example, it is common to claim that students' ability to write can adequately be assessed through their ability to edit already-written pieces, placing the major emphasis on conventions such as grammar and spelling. Teachers who also emphasize the development of more complex skills such as clarity of writing, attention to audience, vibrant language, revision, and sound support of assertions find such tests invalid because they do not reflect the full range of the curriculum.

The processes of reading and writing are complex ones, not measurable by simplistic multiple-choice examinations. For students who are encouraged to adjust their reading to various types of materials and purposes, the narrow selections and tasks that appear on most tests where students must choose a single best answer do not fairly assess the full range of reading ability. In classrooms where students are urged to form opinions and back them up, they need to be assessed accordingly, instead of with tests that do not allow for creative or divergent thinking. Similarly, after being encouraged to deal with entire texts, students do not find that the short excerpts on many tests have much in common with the kinds of reading they have done in and out of class. In addition, such tests send a message to teachers and students about the kinds of reading and writing that are valued by society.

Sound educational practices do not result from a model in which assessment determines curriculum and instruction. Indeed, the more invasive

the assessment, the less valuable the information. Rather, the model must start with a curriculum that includes valuable assessments. Furthermore, the higher the stakes of the assessment, the more important it is that assessments reflect the full complexity of the reading and writing curriculum. Those who have adopted sound curricula, provided the best instruction, and worked to establish literacy as a value in their communities must have access to assessments that support those elements rather than undermine them.

Implications

Interactions among curriculum, instruction, and assessment must lead to productive, ongoing conversations for review and revision. These conversations must include students, teachers, administrators, school boards, parents, test makers, and legislators, all of whom have a stake in valid assessment procedures that do not intrude on the processes of teaching and learning.

Policy makers no less than teachers and students must understand the complexities and importance of a full and critical literacy. They must recognize that tests, although sometimes necessary, are often not the best assessment procedures. They must recognize test results for what they hide as well as for what they reveal. In the public interest, they must not endow test scores with the power to tell more than they are able.

Following is a case study of one school that met the need for accountability by making assessment a part of its ongoing teaching and learning activities.

Instruction Informs Assessment

In one Vermont school system, the school board was demanding more frequent testing. Normative achievement tests were not administered until the end of third grade, but the board believed they needed information earlier in order to tell how well their students were doing in reading during the primary years.

The teachers were experienced professionals opposed to any norm-referenced testing. They and the board arrived at a compromise. The board would wait for two years and not require any further norm-referenced testing. The teachers, during that time, would work toward an evaluation system that more fully informed parents and the board about the instruction and achievement of students in the language arts.

The faculty of the multi-aged (K-3) primary unit had years of experience in teaching language arts in the individualized, literature-based reading program that had been in place in their schools for more than a decade. They wanted to be able to use books to evaluate students and they wanted the assessment to reflect their instructional program. Additionally, they wanted to be able to change an assessment session into instruction if the student's

selection proved too difficult; and they did not want to remove the element of student choice from the assessment process. According to one teacher, "The ability to choose a book at an appropriate level is an important part of being a reader," and they wanted to assess this ability as well as other aspects of reading performance.

The teachers decided to use the computer to enter more than two hundred children's books in the "books for assessment" system. The books were the titles contained in the classroom libraries of these teachers. The computer file for each book would specify its grade-level difficulty (arrived at through discussion and consensus) and contain an assessment summary sheet to be used during evaluation. The sheet provided room for recording the student's retelling of the book, space for noting the six or eight follow-up questions that the teacher asked, along with the answers the students gave, and a list of words from the book to be used for evaluating word recognition.

Either a student or a teacher could initiate an assessment session. In either case, the student would choose a book from those available in the classroom and read it silently. After the student had read silently, the teacher would listen to him or her read aloud, using the assessment summary sheet to evaluate book choice, word recognition, and fluency. The teacher would then ask the student to retell the book, and follow up the retelling with six to eight questions. Finally, the student's isolated word recognition would be evaluated. These assessment sessions would be conducted in very much the same way as the individual conferences that were the mainstay of the instructional program and took place at least twice a year. When necessary, students would be referred for additional assessment by resource professionals who would conduct a more thorough and formal reading inventory across multiple levels.

In their *informed* approach to assessment, these teachers learned more about the processes of reading and writing. They also improved their ability to speak with clarity about their goals and the progress of their students. They felt more confident that they could speak to parents and school board members about important aspects of their curriculum. Further, they became more confident that no individual student would "slip through the cracks." Finally, they knew that they had developed an approach to assessment which was consistent with their philosophy and curriculum, and which also responded to the need to be accountable to others with vested interests in their work.

4. Assessments must recognize and reflect the intellectually and socially complex nature of reading and writing and the important roles of school, home, and society in literacy development.

Rationale

Reading is a complex negotiation among reader, text (an unseen author), purpose, and context. Writing involves similarly complex negotiations

among writer, audience, purpose, and context. For assessment purposes, it is essential to understand that the multiple influences involved in both reading and writing mean that students' behavior in one setting may not be at all representative of their behavior in another. Consequently, reading and writing cannot be assessed as isolated, independent tasks or events. It is critical to gather specific information about materials and tasks being used with students for both instructional and assessment purposes.

In school settings instruction and assessment should be seen as highly interactive processes. For example, aspects of the learning situation interact with cultural and home environments to influence student learning and motivation. These social situations shape purposes for both teachers and students. They also influence the conditions, constraints, and standards present in the learning context and affect students' motivation to engage in reading and writing activities. In the social context of schooling, there are many features known to influence learning and performance. These include types of activities, management efficiency, grouping patterns, teacher and student expectations and beliefs, classroom interactions, and the classroom environment. In addition, factors associated with teaching such as content, tasks, and materials all affect literacy learning.

The quality and appropriateness of assessment efforts depend to a considerable extent on the degree to which these complexities have been considered. The quality of an assessment will be weak if it yields an incomplete or distorted picture of a student's literacy. Characteristics of the text, the task, the situation, and the purpose can all have an impact on the student's performance.

Implications

Only some aspects of reading and writing will be captured in any given assessment situation. Formal tests need to be considerably more complex than is generally true today. Tests that accommodate multiple responses, different types of texts and tasks, and indicators of attitude and motivation are all essential to a comprehensive view of literacy achievement. Wherever possible, assessments must specify the types of texts, tasks, and situations used for assessment purposes and note whether and when students' performance was improved by variations in text quality, type of task, or situation.

In order to meet this standard, we must depend less on one-shot assessment practices and place more value on ongoing classroom performance, assuming that classroom curricula develop the full complexity of literate learning. Finally, when assessment information is interpreted and reported, descriptive information about the assessment tasks and texts and the instructional situation should be included. Given the complexity of the

tasks involved, reducing reading and writing performance to a letter or number grade is unacceptable.

The case study below illustrates the way one teacher found to make sure that all of the factors affecting a student's performance, both inside and outside the school, were taken into account in assessment.

Assessment Must Be Contextualized

Dawn teaches third and fourth grades at Franken Elementary School, a small rural school with a population of seventy students; a faculty of four classroom teachers, one head teacher, whose duties are primarily administrative, and two helping teachers; and a schoolwide commitment to descriptive evaluation and narrative reporting as its methods of assessment. One of her students, Mary, had been making strong academic progress but had some social adjustment problems. Dawn used her summaries (done twice during the second term) and additional notes on Mary to write a narrative account of her academic work and social adjustment. Dawn had had frequent phone conversations and other informal talks with Mary's parents about many physical complaints Mary made that Dawn believed might have a psychosocial basis, and had also consulted with the head teacher. All of these conversations became part of the report. Later, Dawn met with Mary's parents to report her findings and compare them with their perception of their child's progress. Questions, concerns, agreements, and plans emerged from Dawns's conference with Mary's parents, and were recorded on the conference sheet that would become part of Mary's record. One plan was that Mary and Dawn would keep a dialogue journal about the social issues that often seemed to overwhelm Mary.

Dawn's narrative, descriptive account of Mary is a picture of her growth, development, and experience within the school. It is also a picture of the school's response to Mary. Since information came from all sources—student, teacher, parents, administrator, and other teachers—it is as complete as it can be at that particular time and in that particular place. As these reports increase, a cumulative picture begins to emerge. Mary's evaluation is rooted in who she, her family, and the educators are, as well as where they are. In order to evaluate Mary, the context must be part of the evaluation itself. Thus, contextualized evaluation becomes a constant challenge and perspective for Dawn, the other teachers, and the administrator of Franken.

5. Assessment must be fair and equitable.

Rationale

We live in a multicultural society with a constitution that promises equal rights to all. Our school communities must work to ensure that all students, as different as they are in cultural, ethnic, religious, linguistic, and economic background, receive a fair and equitable education. Assessment plays an important part in ensuring fairness and equity, first because

it is intimately related to teaching and learning, and second because assessment is often used in attempts to provide an impartial way of determining who should and who should not be given access to educational institutions and resources. To be fair, then, assessment must be as free as possible of biases based on ethnic group, gender, nationality, religion, socioeconomic condition, sexual orientation, or physical disability. Furthermore, assessment must help us to confront biases that exist in schooling.

In the past, tests have been relied upon to try to avoid the cultural and personal biases of teachers' judgments. However, just as it is impossible to eliminate bias from teachers, it is also impossible to produce a culturally unbiased test of reading or writing. Language itself involves social conventions that differ from culture to culture. Furthermore, words have different shades of meaning for different cultures, and the variation in life experiences across culturally, economically, and geographically different situations can be quite extreme. Consequently, students will differ enormously in the interpretations they give to the texts they read, the topics they feel comfortable writing about, and the ways they respond to different forms of assessment. Even if a test were able to accommodate language differences, the person administering the test could influence individual performance as well.

The inevitability of bias notwithstanding, when tests must be used, as many biases as possible should be controlled, and multiple perspectives and sources of data should be brought to bear to help balance against one another those biases that will inevitably remain. As far as possible, assessment should be accomplished in a language that will not interfere with the individual's performance. Assessment practices should not devalue cultural differences in dialect. Although all students have the right to learn to read and write in the privileged dialect of English (often called "Standard English"), since it is the language of power, failure to use that dialect should not have negative consequences unless the requirement to use the standard English dialect is specifically stated. Moreover, all students should study and be assessed on the languages and literatures of cultures other than their own.

Unavoidable biases should also be made quite clear and public. However, most biases are part of the perspective we bring from our cultural backgrounds, so we ourselves tend not to notice them. For example, people commonly treat boys and girls differently because of the roles and expectations they bring from their own experiences, but are often unaware that they do so. Because it is difficult to notice our own biases, we must be concerned that the testing industry, reflecting the rest of society, is populated and controlled by people whose perspective represents the dominant culture. The teaching profession is faced with a similar problem.

Because cultural and personal biases are often only noticeable to those whose experiences are different, it is particularly important that multiple perspectives be brought to bear on assessment issues (standard 8). For example, one way to take test bias seriously would be to ensure strong and varied minority group representation in the construction of public tests. In this way, test biases should become apparent, and, once recognized, be easier to reduce. A second important way to help address bias is to make the test available for public examination after it has been given, another practice that might render biases more visible.

Because teachers are the most important agents of assessment (standard 7), unless the biases of their individual judgments, expectations, and consequent teaching practices are addressed directly, students will not receive a fair and equitable education. To ensure fairness, teachers must have a deep understanding of cultural and gender conditioning and its consequences, of the sources of difficulty students can encounter in learning to read and write, and of productive approaches to accommodating difficulty and capitalizing on cultural differences.

Other sources of bias are even less obvious. Assessment practices can be biased against an entire curriculum as well as against individual students. For example, a school community may believe strongly in an integrated curriculum but be required to administer tests that privilege the separation of curriculum into subject domains. Similarly, a minimum competency test the results of which are made public will force attention away from a critical, thoughtful literacy not seen as "basic" by the test developer, particularly because it is not possible to examine it using a test.

Inequities in schooling can also be compounded through inappropriate assessment practices. For example, assessment practices often lead to students' being placed in different curricula with the intention of producing a better match between students and curricula. This leads to a major equity issue. On the one hand, a better instructional match is possible, but on the other, different and perhaps lowered expectations, on the parts of both teachers and students themselves, may result. Once students are assigned to systematically different curricula, uneven access to subsequent experiences and jobs becomes not only a possibility but a probability. IRA and NCTE are opposed to the use of tests to control access to educational or employment opportunities when the tests are unrelated to specific employment (or educational) demands, on the grounds that it is an unfair practice.

Other uses of assessment can also produce inequities. For example, external pressures on the use of such tests often differ across school settings, being particularly common in large cities. Similarly, a common practice in some states is to report in the newspaper the average test scores of students by district, school, or even classroom. This practice can lead

to even greater economic differences among districts, since individuals and businesses are reluctant to move into areas whose schools have low scores on the tests. Thus, economically stressed school systems become more so. This practice not only violates standards 1, 2, 4, and 6, but it results in greater inequities.

Although assessment plays an important role in ensuring fairness and equity, the goal of equity cannot be laid solely at the feet of assessment. First, it is not uncommon for school districts to differ in their per-pupil expenditures by a factor of ten. No assessment practice can shore up the differences in educational experience that arise from the obviously unequal conditions of extreme poverty and wealth. Second, otherwise valuable assessment instruments can be used in ways that exaggerate inequities.

Implications

Since the primary assessment agent is the teacher (standard 7), school communities and training institutions must organize to expand teachers' cultural experiences and awareness. School communities must actively work to reduce bias in school assessment practices, much of which will require making people aware of the biases they hold but do not notice, and organizing to bring multiple cultural perspectives to bear on assessment issues.

Performance differences among groups must be used only to improve teaching and learning for all students, not to display for public consumption, and thereby reinforce, already existing stereotypes about language competencies among various groups of students. Furthermore, provisions should be made to ensure that second-language learners be assessed in ways that permit them to show what they know and can do.

There will always be tensions among concerns for individuals, for local communities, and for society at large. For example, some school communities are relatively homogeneous. Community involvement in curriculum and assessment may produce changes in school community practices that are unfair to underrepresented groups in that community. The equity concerns of the larger society would then be in conflict with local concerns. On the other hand, when curriculum and assessment are determined by state or national interests, the motivation for local involvement in schooling is substantially reduced. National, state, and local differences in values and beliefs about schooling, literacy learning, and diversity will produce continuous tensions over curriculum control, involvement, and financial interest. There are no simple resolutions to these dilemmas. They go to the heart of the democratic process. These tensions, then, need to be a source of continual dialogue among the participants in schooling.

6. The consequences of an assessment procedure are the first, and most important, consideration in establishing the validity of the assessment.

Rationale

What assessment improves determines its value. No matter what other attributes an assessment procedure has, its consequences for students are primary. Any assessment procedure that does not contribute positively to teaching and learning should not be used.

By asserting that procedures cannot be evaluated out of the context of their use, this standard puts assessment, teaching, and learning back together. It asserts that simply devising a more detailed or more complex test will not directly result in a more valid assessment. If an assessment procedure has adverse motivational consequences for school communities, or segments of school communities, then the procedure is invalid. Adverse consequences from assessment can arise from a variety of procedures, for example:

a. Assessment techniques that very publicly value only a narrow range of literacy activity, or very controlling forms of reading and writing, and thus enforce a narrowing of the curriculum for students.
b. Assessments or reporting procedures that make students or teachers defensive or unreflective, thus inhibiting learning.
c. Reporting procedures that focus on ranking or rating rather than on performance, thus drawing learners' attention away from the process of learning, reducing learners' notions of literacy to a simple linear continuum, and disrupting collaborative learning communities.
d. Assessments that oversimplify complex literacy behaviors or situations and consequently narrow the curriculum, for example when classroom assessment focuses on worksheets and multiple-choice tests, or when evaluative feedback to students' writing focuses only on spelling and grammar and not on students' thinking, substantive content, or organization.

In its basic form, this standard is not new to most people in the field of assessment. However, in the past it has largely been ignored. It is commonplace to talk about different purposes for assessment and to invoke the principle that the assessment must match the purpose for which it is intended. Nonetheless, test publishers have taken no responsibility for the use of their tests, while asserting their validity. This standard denies the unfortunately common argument that a given test was valid in

spite of the fact that it was used to place a student in a program that does not serve her well.

Implications

This standard essentially argues for "environmental impact" projections and piloting prior to any institutional use of an assessment procedure, along with careful, ongoing analyses of the consequences of its use. The burden of this standard is on the entire school community, to ensure that assessments are not used in ways that have negative consequences for schools and students. This standard means that assessment information should not be used for judgmental or political purposes if that use would be likely to cause harm to students or to the effectiveness of teachers or schools. For example, newspapers have the responsibility not to report school assessment data without putting them into context—including the availability and quality of resources, teacher and student profiles, and location—nor to report them in a manner that will have destructive consequences, such as making comparisons among schools or among teachers. Similarly, schools have a responsibility to report student assessments to parents in a way that will assist students' learning and parents' understanding, not hinder them.

Producers and developers of assessment instruments must take responsibility for the instruments they develop. In light of what we have learned about the iconic or symbolic value of tests in shaping the curricular decisions made about students by teachers, administrators, and policy makers, a "user beware" attitude is unacceptable within the framework of this standard. For example, a multiple-choice, machine-scored test used to hold teachers accountable will almost certainly have adverse effects on the curriculum that students experience in the classroom. If assessments are to be used for high-stakes purposes such as setting public policy or holding people accountable, then they must be fully consistent with, and not a shorthand for, the assessment procedures used to provide teacher and student with knowledge of progress in the classroom. They must be consistent with standard 3 and reflect the curriculum.

Environmental impact reports are never simple issues and should be ongoing. The important thing is to ensure careful consideration of consequences, capitalizing on multiple data sources and multiple perspectives, always recognizing that these efforts are likely to raise value-laden conflicts such as the tension between the public's right to know and the preservation of conditions that will foster learning.

7. The teacher is the most important agent of assessment.

Rationale

Assessment instruments have traditionally been conceived of as tests, and teachers have been viewed as merely consumers of information generated by these tools. However, of all the evaluation that takes place in education, most takes place in the classroom, as teachers and students interact with one another. Teachers design, assign, observe, collaborate in, and interpret the work of students in their classrooms. They assign meaning to interactions and evaluate the information that they receive and create in these settings; in short, they do function as agents of assessment, and their assessments have enormous impact on students' lives.

Although teachers have functioned as principal evaluators for a long time, they have rarely been thought of as assessment agents. This standard, then, in some ways simply acknowledges formally a critical role of the teacher and the consequences and responsibilities that accompany this role. In fact, teachers bear the responsibilities implied by each of the other standards in this document.

Whether they use tests, everyday work samples, discussion, or observation, teachers make sense of students' reading and writing development. To use the metaphor in the introduction, they "read" the many assessment texts that students provide in order to "construct" a reading of students as literate individuals. The sense they make of a student's reading or writing is passed on to the student through spoken or written comments and by instructional decisions made in the classroom (e.g., subsequent assignments, grouping). Teachers already engage in these types of activities; however, this standard asserts that they must be deliberate about and aware of their roles as assessors.

In the past, this society has valued and invested in external assessment practices such as testing, and devalued the information that teachers gather daily about their students' learning. As a result, some teachers have even learned to devalue their own knowledge. But teachers are in a unique position to engage in valid assessment. Because they are the closest to students' learning, they have the opportunity to make many detailed observations over time. For example, the use of developmental portfolios can reduce the likelihood that a student's "bad day" performance will unduly influence a teacher's conclusions about that student's overall performance. It also allows a wider range of observations to be made in a more diverse and representative set of situations, thus increasing the validity of the assessments that take place. The teachers' situation also allows them to adapt assessment to the special characteristics of individual students, instructional programs, and community expectations, as well as

making it likely that instructional practice will be affected by the assessment.

Teachers' observations are certainly open to concerns about bias and lack of objectivity, an objectivity that published tests superficially appear to offer. And teachers' assessments have many more consequences than external testing. Even if it were possible to construct an unbiased test of literacy (and our understanding of language asserts that it is not), if the teachers' assessment is limited and biased, then the consequences for the students remain unsatisfactory.

The basis for less-biased assessment repertoires—teachers' classroom assessment practices as well as their use of other assessment tools—is teachers' knowledge about learning and literacy. The more teachers know about literacy development in general and, more important, about the literacy development of individual students, the more they will be able to make sense of what students do and the better equipped they will be to provide appropriate instruction. The foundation of this assessment ability is deep and diverse knowledge of individual students and of reading and writing.

This knowledge cannot be replaced by tests of any kind. First, any one-shot assessment procedure cannot capture the depth and breadth of information teachers must have available to them. Even when a specific (perhaps multiple-choice) test is used, teachers must use the full range of their knowledge about content and individual students to make sense of the limited information provided by such a test. Second, a teacher who knows a great deal about the range of techniques readers and writers use will be able to provide students (and parents) with specific, focused assessments. Third, students learn different things about themselves and about literacy from teachers' evaluative feedback than from standardized tests. If teachers' comments are specific, providing a clear picture of each students' special strengths and weaknesses, rather than comparing students to each other, these characteristics will be reflected in students' self-evaluations. Similarly, if a teacher enables students to engage in self-evaluation, then students will be encouraged to take control of their own literate learning.

Implications

First, for teachers to be effective agents of assessment, they must have a deep knowledge of the disciplines of reading and writing and an understanding of their own reading and writing. Therefore it is imperative that teachers read and write themselves and discuss their reading and writing with others. They also must cultivate an ongoing understanding of students' reading and writing development, of instructional strategies, and

of assessment techniques. The more knowledgeable teachers are about reading and writing, and the more observant they are of students' literacy behavior, the more productive will be their assessments and their interactions with students.

Second, teachers cannot be expected to acquire and refine this knowledge without considerable support. Indeed, the major investment required for improving assessment must be in staff development and school-community learning. Serious attention must be given to providing the time and conditions that will help teachers maximize and reflect on their knowledge. Procedures that provide credible data to teachers in situations that encourage discussion and involvement, and reduce defensiveness, are likely to be most productive. Such conditions encourage the engagement of the multiple perspectives necessary both for learning and for reducing the effects of individual biases. Group-administered, machine-scored tests, on the other hand, do not stimulate reflective teaching and learning. Neither do conditions of external accountability. Consequently, these are not productive ways of improving literacy instruction.

Third, when teachers serve as agents of assessment, they must take responsibility for making and sharing judgments about students' achievements and progress. They cannot defer to others or to other instruments. At the same time, others must come to trust and support teachers in their judgments. Such trust and support are fostered when school communities are organized in ways that bring multiple perspectives to the assessment process and counter any inherent bias (see standard 5).

8. The assessment process should involve multiple perspectives and sources of data.

Rationale

Perfect assessments and perfect assessors do not exist. Every person involved in assessment is limited in his or her interpretation of the teaching and learning of reading and writing. Similarly, each text and each assessment procedure has its own limitations and biases. Although we cannot totally eliminate these biases and limitations from people or tests, we can try to ensure that they are held in balance and brought to awareness. The more consequential the decision, the more important it is to seek diverse perspectives and independent sources of data. For example, placement in, or eligibility for, specialized programs has a profound influence on a student's life and learning. Such decisions are simply too important to make on the basis of a single measure, evaluation tool, or perspective.

The need for multiple indicators is particularly important in assessing reading and writing because of the complex nature of literacy and its acquisition. A single measure is likely to be misleading or erroneous for individuals or groups. Multiple sources of data, on the other hand, can allow for "triangulation" in problem solving. Sources of data can include, for example, observations made in different situations or by different people at different times, or data from different assessment instruments. However, data from more than one of the same kind of assessment instrument (for example, a series of standardized tests) will not satisfy this standard because such tests commonly reflect a similar, and narrow, view of literacy. By the same token, even new data can be looked at with old eyes. Unless different perspectives and values are brought to bear on data, our understanding might not expand. Even the richest set of data can be reduced to mere conventions by a limited perspective.

From a statistical point of view, the reliability of assessment interpretations is likely to improve when there are multiple opportunities to observe reading and writing. Adherence to this standard will also substantially improve the validity of the literacy assessment process because sampling more than one aspect of literacy permits a closer approximation of the complexity of reading, writing, listening, and speaking processes as they occur and as they are used in real-life settings.

However, seeking multiple perspectives and sources of data is not simply a matter of reducing biases or errors in individual data sources. It is, rather, taking advantage of the depth of understanding that varied assessment perspectives afford and the dialogue and learning they produce. For example, two teachers with different cultural backgrounds might interpret a student's literacy development in different but important ways. Indeed, because literacy learning is also social in nature, these two teachers' different interpretations will lead to different kinds of development. The exploration of these contrasting perspectives will lead not only to a more productive understanding of the specific student's development, but also to an enhanced awareness of possible interpretations of other students' development, and of what it means to "develop."

Implications

School communities have the responsibility to develop an atmosphere in which diverse perspectives on learning are valued and engaged in dialogue as part of decision-making. This will also form the basis of community learning. Teachers have a particular responsibility to engage in such dialogue both as a way of expanding the range of metaphors they have available for understanding students' literacy development, and as a way of revealing to themselves their own instructional values—how they

have construed development. Multiple perspectives provide the tensions necessary for promoting dialogue and inquiry.

Attempts to satisfy this standard require some cautions. For example, multiple assessments are unlikely to improve the assessment process when they are externally mandated without adequate reference to the context of the school, the classroom, and the needs of the students. Insofar as stakeholders in education know and make known the extent to which given learning communities differ with respect to location, quality and availability of resources, and faculty and student body profiles, the more or less meaning multiple assessments will have in the assessment process. Furthermore, the process of inquiry will be severely limited when particular perspectives or sources of data hold privileged status. Currently, for example, standardized tests hold such status, which reduces the possibility for productive dialogue. Similarly, certain cultural perspectives are given privileged status in assessment. School communities must work to overcome these pressures. They must draw on the larger context of home, culture, and community in order to expand their perspectives and their sources of data.

Multiple perspectives and data sources should be built into our classroom and school-community routines, but we must take care that continuous examination of multiple sources and perspectives provide teachers and students with additional benefits rather than additional burdens.

The following case study offers an example of an assessment that drew on multiple perspectives and sources of data in order to enhance the learning of a student who had been labeled as learning disabled using norm-referenced standardized tests.

Special Education, People Instead of Tests, and John

Mario Jordan—young, energetic, and absolutely horrified about John's situation—was on temporary loan to Forrest School from another school in the district. John had been on an Individual Educational Plan for three years. He had been found eligible for special education services under the learning disabled category at the end of his second year in first grade. He had been tested using the following instruments: BSSI, TERA, Woodcock Johnson Psycho-Educational Battery, McCarthy Scales of Children's Abilities, WISC-R, Piers-Harris Children's Self-Concept Scale, Social Competence Checklist, Family Functioning Survey, and a clinical interview. On all measures of intellectual functioning, John's scores fell in the average or even high average range. His academic achievement scores, however, were extremely low (both reading and written language scores fell below the first percentile).

Now three years older and in fourth grade, John had made painfully small gains. According to Mario and the classroom teacher, he was still reading at only the pre-primer level. Mario called Susan, at a nearby university's

reading clinic, and asked about the possibility of John's attending the clinic. Reluctant to accept children into the clinic who had already been found eligible for special education services, Susan nonetheless relented and decided to accept John into the clinic *on condition that the information the clinic gathered and the recommendations made would be entered into the reevaluation record.* In other words, the clinic's work would be considered assessment, even though it did not expect to administer any norm-referenced tests.

The first priority was to gather additional information in order to answer the questions that earlier evaluations had not clearly addressed: What *could* John do? What *did* he know? The file was silent on these questions. There was a great deal to know about what John could not do and what he did not know, but little about his knowledge and skill in reading and writing. The earliest interactions between the clinic and John, however, painted an even bleaker picture than the record showed. After five years of formal schooling, John could read fewer than forty words (the number varied across isolated and contextual reading tasks but at no time could the clinic collect a corpus that suggested he could read even fifty words). He had difficulty rereading his own dictated texts and he was able to write only about ten words unaided. More troubling, given the heavy focus on phonics he had received for four years, he was unable to segment sounds sufficiently to use sound-symbol knowledge in invented spelling. Attempts at using sound-symbol analysis during decoding were slow and largely unsuccessful, and he could reread his own dictated sentences only during the short term, appearing to rely heavily on memory versus print cues.

What would reestablish learning for John? The clinic decided to use a language-experience approach and related activities to teach reading, which John's special education teacher and psychologist had recommended, but which the classroom teacher in reading had been unwilling to use, instead using a synthetic phonics approach to sound-symbol analysis. After three months of weekly instruction in the clinic, John had made progress. He was able to sustain himself during silent reading for five minutes, completing a hundred-word portion of a Dr. Seuss book. Using a repeated reading approach, he was able to read seventy-five to eighty words correctly in primer-level materials. John was enormously successful in using a computer and word-processing program with voice synthesizer for writing, planning, and completing a "picture" book based on the book *I Want a Dog* by Dayal Khalsa. The word-recognition skills he built during these activities resulted in a reliable word recognition corpus of better than seventy-five words.

However, it was now time for John's reevaluation. Instead of repeating the administration of norm-referenced tests, Mario sought and received permission to use John's tutor and Susan as evaluation agents. They provided copies of John's final clinic report, and Susan appeared before the support team with the tutor's and her appraisal of John's progress and continuing needs. Their recommendations for instruction—both the content and approach—were supported by John's parents and the special educator. Thus, the teacher became the agent of assessment. Not only did Susan and the tutor analyze and evaluate what progress John had made, but made recommendations for continuing instruction, supported by the special educator and John's parents.

9. Assessment must be based in the school community.

Rationale

There are many possible foundations on which to base assessment, from the classroom to the district or state. We believe that the appropriate foundation is the school community.

Assessment should involve the community outside the classroom, for several reasons. First, those members of the school community outside of the schools have a stake in assessment because of their stake in the lives of their children and in the future of society. Assessment is always value-laden, and the ongoing participation of all parties is necessary in a democratic society. Second, the collective experience and values of the community can offer a sounding board for innovation and offer multiple perspectives on a situation to provide depth of understanding. Third, because language learning is not restricted to what occurs in school, curriculum inquiry requires that assessment too go beyond merely the school curriculum. Fourth, the involvement of all parties in assessment encourages a cooperative, caring relationship among them rather than an adversarial one. On the other hand, the school community is also a more appropriate foundation for assessment than any larger unit such as the school district, county, or state. When inquiry is based in such larger units, the distance from the problems to be solved and among the participants reduces the probability of feelings of involvement and commitment and increases the possibility that assessment will become merely a means of placing blame.

With the school community as a center of inquiry, diversity of perspective is possible, not only as a source of growth for individual classroom teachers but also among teachers and the larger school community. Diversity of perspective is what brings depth of understanding and productive problem solving, and face-to-face involvement brings personal knowledge of the issues of assessment as well as personal investment in them. If teachers are able to make informed assessments and articulate them well, it is largely because they have engaged in dialogue about their students' reading, writing, and learning, and been supported by the larger community in doing so. In order for a school community to do this effectively, it is necessary to engage in self-examination and make learning within the community a priority.

Implications

To function as a center of inquiry, a school must develop a trusting relationship with its community. This relationship commonly grows out

of involving all members of the community, balancing power, and recognizing different points of view. Because building such a relationship is nearly impossible in the context of large schools, whose hierarchical structures discourage the openness necessary for reflection, discussion, and inquiry, manageable schools-within-schools become an important possibility to be considered.

There must be an ethos that educators are learners, too, particularly about their own role in students' learning and the operation of their institutions. In order for educators to learn from others' perspectives, school communities bear particular responsibility for ensuring that all their members become fully involved in the assessment process. Many parents, for example, partly because of cultural disparities or their own schooling histories, do not feel comfortable voicing their concerns. School communities have a responsibility to create conditions and assessment procedures that make people comfortable doing so.

As parents become more fully involved in schools and assessments, they become more informed about and more observant of their children's development. This involvement allows them to be more supportive of their children's learning and the teachers' efforts, and leads them to articulate more clearly their concerns about their children's progress. Furthermore, when administrators, parents, and the public become involved together in school assessment issues, trusting relationships are likely to evolve. With a trusting relationship, members of the school community can confront limitations and weaknesses as well as recognize strengths of their curriculum and assessments.

If the school curriculum produces a highly controlling model of literacy, or wittingly or unwittingly invokes cultural, gender, or other stereotypes, the community should be made aware of such curricular problems through assessment practices. It is never easy to overcome the restrictions placed on approaches to assessment by the unquestioned assumptions that members bring from their own educational histories; however, community diversity and exploration of other communities' solutions can help. Since the teacher is the central agent of assessment, the teaching community is responsible for maintaining communication about the process of classroom assessment with other interested parties.

The following case study illustrates the way one school community participated together in assessment.

All Stakeholders Must Participate in Assessment

Franken Elementary School is small and rural, with a population of seventy students and a faculty of four classroom teachers, one head teacher, whose duties are primarily administrative, and two helping teachers. It is

committed to descriptive evaluation and narrative reporting because the faculty believes those are the best ways to judge the curriculum and the students' participation in it. Students engage in reading books, writing for real purposes, and computing, measuring, and working with numbers to solve actual problems, rather than engage in contrived activities. Many of the skills and concepts that the students develop in reading, writing, and computing are applied and enhanced in their study of content-area subjects, which are organized as topics under broad themes such as "Explorations." The students continually produce projects and performances which show what they know and what they are able to do. On a daily basis the teachers record their observations of the students' processes—*how* they are learning—and progress—*what* they are learning. They also record notes on conferences, which are conversations between the teacher and student(s), or among students, about what their understandings are. By their very nature many of these conferences have components of evaluation within them. Evaluative comments ("I read this whole book by myself") and questions are part of the conversation.

Some conferences are strictly about evaluation. At that time students respond to questions, such as "What is the most important thing that you have learned in math since the beginning of the year?" or "What can you do now as a reader and writer that you could not do in November?" These conferences with students are scheduled near the end of each term (November, March, and June); the teachers also schedule conferences at the end of the first and second terms (November and March) with parents, but at the end of the third term (and school year), the teachers write reports to the parents, analyzing the patterns and insights that they see emerging in their children.

Evaluation time at Franken Elementary School is a busy time for students, teachers, administrator, and parents.

10. All members of the educational community—students, parents, teachers, administrators, policy makers, and the public—must have a voice in the development, interpretation, and reporting of assessment.

Rationale

Each of the constituents named in this standard clearly has a stake in assessment. Students are concerned because, among other things, their literacy learning, their concepts of themselves as literate people, and in many ways the quality of their subsequent lives and careers are at stake. Teachers have at stake their understandings of their students, their professional practice and knowledge, their perceptions of themselves as teachers, and the quality of their worklife and standing in the community. Parents clearly have an investment in their children's learning, well-being, and

educational future. The public invests money in education, in part as an investment in the future, and has a stake in maintaining the quality of that investment. The stewardship of this investment involves administrators and policy makers.

Following the constitution, policy makers have the responsibility of ensuring equity and preventing local injustices.

However, when policy makers develop assessment practices that drive local assessment and instructional processes, other stakeholders' voices are easily silenced, and assessment becomes dominated by procedures developed by people with little regular contact with students or teachers. The stakes are too high for this situation, which too often currently prevails, to continue. The further away from students they are, the less informed people tend to be about the students' situation and the less able they are to make informed sense of assessment data. In addition, the greater their distance from the site of assessment, the less personal involvement people have. At any rate, as we have argued, the development of assessment practices is too important to be left in the hands of any one individual or group.

Combining the perspectives of different participants can prove most productive. On the one hand, distance can be seen as an advantage in that it can allow a view of the big picture, and maintain a concern for equity that reaches beyond the local community. On the other, distance tends to blind decision makers to the significant consequences assessment can have for individuals and communities. However, in order for productive use of multiple perspectives to emerge power differences among participants must be equalized.

It is common, for example, for students and teachers to feel that assessment is something that is done to them rather than something in which they are involved. A common reaction to this feeling is to reject the value and credibility of the assessment procedure. At the same time, there is a breakdown in the relationship between those controlling the assessment and those who feel controlled by it. But the more ownership the various participants feel in the assessment process, and the more seriously they value their own and others' stake in the process, the greater the possibility of quality assessment.

Implications

Because the different stakeholders have different concerns, they will have different notions about what constitutes the most effective approach to assessment. Consequently, the public's interest, the parents' interests, the teachers' interests, and the students' interests must all be represented in a way that does not privilege particular groups. This means, for ex-

ample, that policy makers should not decide on a particular course of action and then allow teachers to participate only in finalizing an assessment procedure in a pretense at involvement. Neither should teachers or schools ignore the interests of parents or students in the development of assessment procedures. They must be involved from the beginning. Development of assessments should be continuous in order to accommodate and encourage the involvement, and changing concerns, of all parties.

Assessments are ways of viewing and reporting what is happening in schools, and teachers and other educators need to be clear about what they are reporting. The law mandates attendance, and parents and children have little or no choice of schools and programs. Thus, teachers have a special responsibility within the school community to explain clearly, and in detail, what students are doing in school and to respond clearly to parent and student inquiries about learning. This responsibility must be supported by the entire school community.

11. Parents must be involved as active, essential participants in the assessment process.

Rationale

The current state of public education is characterized by unequal funding resources between one school district and another and by unequal participation of parents in all aspects of school activities. The first characteristic is chiefly responsible for the unevenness among school districts in facilities, resources, quality teaching, sound learning, and healthy environments conducive to effective teaching and learning. The second condition, unequal parental participation, contributes significantly to the difference between productive and unproductive schools. Arguably, the best schools have highly active participation by parents in all aspects of governance and activities.

Teachers and students have the largest stake in curriculum, instruction, learning, and assessment. However, parents also have an important stake in, and responsibility for, their children's education. In many schools, parents stand on the periphery of the school community, some feeling hopeless, helpless, and unwanted. Parents must become active participants in the assessment process. Parental involvement in assessment, which is inseparable from curriculum, instruction, and learning, includes the following:

a. Becoming knowledgeable about assessment. Because of their own schooling backgrounds, many parents believe that report-card grades

and test results from multiple-choice examinations are the most productive and informative measures of their children's performance, knowledge base, and achievement. They need to become knowledgeable about the diverse possibilities for assessment, what those possibilities have to offer for understanding and assisting their child's development, and the uses and misuses of various forms of assessment.

b. Actively participating in the assessment process and all other aspects of governance in their school community.
c. Contributing their knowledge of their children. As an important part of all school communities, parents have valuable knowledge of their children's development and situations that can contribute to the assessment process.
d. Seeking ways to become more knowledgeable about their children's development.

Paying school taxes does not absolve parents from the responsiblity for participation in their children's education any more than paying taxes absolves them from the responsiblity for voting. Teachers need the knowledge parents have of their children, and school communities need the diversity of perspective that parents bring to school problem solving, including assessment.

Implications

The responsibility for parental involvement lies on both parents and schools. Parents must seek ways to become involved, and schools must organize to include parents in their assessment and staff development programs, and actively seek their participation. This is particularly important in the case of parents who are frequently marginalized by society in general, and by the school system in particular.

The size and nature of the school community will have an impact on the ease with which parents will be involved in schools, and the resources necessary to increase their participation. Consequently, this standard implies adequate and equitable funding of schools.

Involving parents in the assessment process includes involving them in staff development or community learning projects in which they learn more about reading and writing. It also includes the use of communication and reporting procedures between school and home that enable parents to talk in productive ways with their children about their reading and writing. Involving parents and parent committees in the development of new reporting procedures is essential, since they are the primary audience for such reports.

Glossary of Assessment Terminology

Rapid changes in the field of reading and writing assessment have generated a variety of new terms as well as new uses for many established terms. The purpose of this glossary is to specify how assessment terms are generally used in discussions of literacy assessment and to point out alternative meanings of terms where they are common. We begin with *curriculum* since it is the foundation for our understanding of assessment as curriculum inquiry.

Curriculum

We can think of curriculum as having three components: the envisioned curriculum, the enacted curriculum, and the experienced curriculum. The envisioned curriculum is the vision of the lives we would like to live in classrooms. The enacted curriculum is our daily attempts in classrooms to put the envisioned curriculum into practice. The experienced curriculum is the sense the language learner makes of what goes on in the classroom and is thus constructed within the language of that classroom. For example, if most of the reading material in one class involves racial or gender stereotypes, then that is likely to be reflected in students' learning, and, by contrast, students are likely to construct different knowledge about human relationships from a more balanced selection of reading material. However, the knowledge and attitudes students construct from those works is strongly influenced by the ways the teacher talks about them, the nature of group discussions, and the ways teachers and other students respond to each other. Ultimately, it is the experienced curriculum that is our concern, and that is why students must be our primary curricular informants. However, it is the discrepancies among the envisioned, enacted, and experienced curricula that drive curriculum inquiry, the process of assessment.

Standards 1, 3, and 4 are particularly closely related to issues of curriculum.

Aggregation

In assessment, aggregation is the process of collecting data together for the purpose of making a more general statement. For example, it is

common practice for school districts to add together all of the test scores for their students in order to find the average performance of students in the district. This process strips away all of the differences among the various cultural groups, schools, and students within the district in order to make the larger statement. Even an individual student's test score is a result of aggregating all of the individual items to which the student responded in order to make a general statement about a student's "reading ability." It is also common then to "disaggregate" the scores to see how subgroups performed within the larger group. Aggregation and disaggregation are in some ways a matter of deciding what are relevant and what are irrelevant data.

There are powerful tensions in this society around the issue of aggregation—reflecting, on the one hand, the need to make general statements about students, teachers, and schools, and, on the other, the problem of stripping away the particulars of individual performances and situations in the process. It is not universally agreed that it is valuable to reduce students or schools to numbers, let alone for which purposes or on what grounds that might be reasonable. It is often argued that administrators need highly aggregated data to make programmatic and budgetary decisions. However, both in education and in industry administrators make different decisions when facing aggregated data than they do when facing real situations with real persons.

Authentic, Performance-based Assessment

These terms and the kinds of assessment to which they refer arise from the realization that widely employed assessment tools generally have been poor reflections of what literate people actually do when they read, write, and speak. The logic of authentic assessment suggests, for example, that merely identifying grammatical elements or proofreading for potential flaws is not an acceptable measure of writing ability. For their writing to be assessed, students must write, facing the real challenges faced by literate people.

The general issue of the "realness" of what is being measured (its construct-validity) is alluded to by the terms: *authentic assessment, performance-based assessment, performance assessment,* and *demonstrations.* Regardless of what the assessments are called, the issue is that tests must measure what they purport to measure: a reading test requires a demonstration of, among other things, constructing meaning from written text; a writing assessment requires a demonstration of producing written text.

Controversy continues to exist about whether machine-scorable, multiple-choice tests have a place in a world in which the criterion of authenticity is applied systematically and rigorously to the evaluation of

assessments. The issues of authentic, performance-based assessment are particularly relevant to standards 4 and 6.

Equity

Issues of fairness surround literacy assessment. Testing originated as a means to control nepotism in job selection—providing an independent perspective on selection to uphold fairness. But equity cannot be assured through testing alone. Those who control the assessment process control what counts, or what is valued. As we pointed out in the introduction, language assessment is laden with cultural issues and biases. Although equity cannot be assured through assessment, it must be pursued relentlessly in assessment and in schooling. It is more likely to be achieved through the involvement of multiple, independent perspectives than through the use of a single perspective.

Tests have traditionally been administered, their results published, and their impact on instruction instigated with little regard to issues such as cultural, economic, and gender equity. But many equity issues affect assessment, rendering comparisons difficult and often meaningless. Because traditional test makers have all too frequently designed assessment tools reflecting narrow cultural values, students and schools with different backgrounds and concerns often have not been fairly assessed.

Equity issues also include the kinds of educational experiences available to students who will face similar assessments, particularly in certification or gatekeeping situations. Questions of access to sound instruction, appropriate materials, and enriching learning opportunities are critical. Educators have become increasingly aware of the connections between assessment results and levels of safety, health and welfare support, and physical accessibility. Any responsible assessment must engage the full complexity of situations faced by educational communities. These issues related to equity are most closely tied to standard 5, but touch all the standards here.

Norm-referenced or Criterion-referenced Assessment

"Referencing" is choosing a framework for interpreting something, in this case assessment data. Norm-referenced interpretations are based on comparisons with others, usually resulting in a ranking. A norm-referenced interpretation of a student's writing will assert, for example, that the sample of writing is "as good as that of 20 percent of the students in that grade nationally." Criterion-referenced assessment is based on predetermined criteria that serve as "yardsticks" or "benchmarks" of performance. Neither frame of reference is particularly illuminating instructionally.

Other, less common frames of reference are more productive in that regard. For example, performance can be interpreted in the context of previous performance (self-referenced). Performance can also be interpreted in the context of a particular theory of literate learning (theory-referenced). But these frames of reference have consequences for the whole process of assessment. They bring with them consequential changes in assessment procedures. In order to make self-referenced assessments one needs to arrange for the collection of historical examples. In order to make theory-referenced interpretations one has to have a coherent theory. To make norm-referenced assessments, assessment practices need to be standardized and focus on maximizing the differences among individuals on a single scale.

Norm-referenced testing is the most prevalent form of large-scale testing, in which large groups of students take a test and the scores are grouped and interpreted in relation to other scores. In other words, the score of any student or group (school, district, state, or nation) has meaning only in relationship to all the other scores of like entities, e.g., school to school, district to district, state to state. In order to make such comparisons, we have to make the assumption of "all else being equal." In other words we try to make everything the same so that differences in performance can be attributed to one source: the student, or school, or district—whichever is the level of *aggregation*. This assumption, as we pointed out earlier in our discussion of language, is extremely dubious. It does not usually take into account the differences that abound throughout the thousands of schools and districts relating to curriculum, culture, gender, ethnicity, economic circumstance, per-pupil funding, and so forth. National norm-referenced tests assume that all students in our society have had similar cultural and curricular experiences.

Norm-referenced interpretations often occur in classrooms, too. A teacher who has little knowledge of the complexity of literacy learning will often have to resort to comparisons and rankings in order to interpret students' reading and writing. Such normative assessments often turn up as grades on report cards. Teachers with a reasonably detailed knowledge of their students' reading and writing, on the other hand, will have difficulty reducing their knowledge to simple rankings for such purposes. Indeed, the process poses highly stressful ethical dilemmas for them. Although grades and rankings are a common part of the educational history of most individuals in this culture, this committee believes the practice to be unnecessary and generally counterproductive.

Some of the stakeholders in assessment—parents, teachers, students, administrators, policy makers—have been seduced into believing that norm-referenced test scores are readily interpretable and productive. However, when it comes to assessing reading and writing, norm-referenced

test scores have little utility because they oversimplify highly complex processes. These processes cannot be evaluated by a machine-scored, multiple-choice test—the most common form of norm-referenced assessment. Assessments based on norm-referenced tests give at best inadequate and often actually misleading information about many students. Most unfortunately, norm-referenced test scores have too often become the single most important criterion for decisions about placement and promotion that have a powerful impact on students' lives.

Criterion-referenced testing involves tests that compare students' performance against established benchmarks. These benchmarks or criteria are usually expressed as numerical ranges that define levels of achievement. For example, an 80-85 score may mean high performance among levels of achievement ranging from unsatisfactory to outstanding. Criterion-based testing can also involve holistic scoring of writing, for example, where a score is based on a set of pre-established consensual criteria.

Standards 1, 2, 3, 4, and 6 raise issues related to norm-referenced and criterion-referenced assessments.

Reliability

Broadly speaking, reliability is an index of the extent to which a set of results or interpretations can be generalized across tasks, over time, and among interpreters. In other words, it is a particular kind of generalizability. For example, a common concern raised by newer forms of literacy assessment is whether different examiners, evaluating a complex response and using complex scoring criteria, will draw similar conclusions about a student's performance (whether an assessment will generalize across different examiners). Experience from scoring complex student writing samples does suggest that when people are well trained in the application of specific criteria, high rates of agreement can be achieved; however, this agreement does not guarantee a high-quality assessment. Indeed, current assessment practices stressing reliability as the central quality of assessments generally focus on trivial matters, on which it is easiest to gain agreement. Reliability is only important within the context of validity—the extent to which the assessment leads to useful, meaningful conclusions and consequences.

In order to provide more "authentic" tasks, newer approaches to testing reading use more substantial bodies of text than the brief excerpts typical of older tests. Because these require more reading and response time, fewer assessment tasks or "items" are typical. For example, rather than having students read and answer multiple-choice questions about a dozen or more short passages, students may be asked to read one or two long

pieces. The specific content of those passages may seriously influence that student's performance. This would limit the generalizability of any statements made about the student's reading of expository materials. In "one-shot" tests, there is thus a trade-off between the extent to which one can generalize performance in reading and writing to real ("authentic") situations, and whether one can generalize across examiners or tests.

One way to increase the reliability of statements about students' reading and writing performance while maintaining authenticity is to avoid dependence on one-shot tests, taking more advantage of continuous classroom assessment, at least where classroom practices reflect the literate activities of the real world. Standards 4 and 5 raise issues related to the reliability of assessment.

Validity

Historically, a common definition of a valid measure is that it measures what it purports to measure. The evidence for the validity of most reading and writing assessment tasks in the past was very thin, or nonexistent, often consisting only of how well a new test of reading, for example, correlated with some other measure of reading. If assessments of literate learning are to measure what they purport to measure, they will need to concern themselves with the nature of language. Valid assessments must then respect and value student diversity and acknowledge that there is generally no single "correct" response. Such assessments would allow for and encourage multiple interpretations of a reading selection and make provisions for allowing students to demonstrate their ability to construct meaning through multiple response modes such as writing, drawing, speaking, or performing.

To a very great extent, a valid assessment is one that reflects a valid curriculum. But more recent conceptions of validity include an examination of the consequences of assessment practices. In other words, one cannot have a valid assessment procedure that destroys curriculum in the process. Consequently, a more productive definition of a valid assessment practice would be one that reflects *and supports* a valid curriculum. As standard 6 asserts, assessment must have consequential validity.

Validity issues are addressed particularly in standards 1, 2, 3, 4, 5, and 6.

International Reading Association
800 Barksdale Road, P.O. Box 8139
Newark, Delaware 19714-8139
ISBN 0-87207-674-1

National Council of Teachers of English
1111 West Kenyon Road
Urbana, Illinois 61801-1096
ISBN 0-8141-0213-1

BEST COPY AVAILABLE